图书情报工作
第64卷 第2期 2020年1月

传统节日文化标签系统及隐性知识挖掘*

■ 苏瑞竹[1] 闫静雅[1] 欧阳剑[2]
[1] 广西民族大学管理学院 南宁 530006 [2] 上海外国语大学图书馆 上海 201620

摘 要：[目的/意义]传统节日文化表达的复杂性，往往难以从中挖掘出更多的深层含义，本研究利用知识组织方法，构建标签系统对传统节日文化知识进行整理与表达，挖掘节日文化背后的隐性知识。[方法/过程]首先设计和定义传统节日文化标签系统的结构与属性，其次爬取并整理了一定网络数据，最后从多个维度挖掘其背后的隐性关联。[结果/结论]提出基于传统节日文化标签系统的隐性知识挖掘方法，实验结果与客观规律相符，证明其具有一定的有效性和可靠性。

关键词：传统节日 标签系统 隐性知识 知识挖掘
分类号：G127 G203
DOI：10.13266/j.issn.0252-3116.2020.02.014

节日文化是一个民族重要的文化遗产，具有丰富的历史文化内涵，集中展示了民众的精神信仰、审美情趣、伦理关系与消费习惯[1]，联合国教科文组织的《保护非物质文化遗产公约》将节庆活动纳入保护范围[2]。传统节日文化是我国民俗文化的重要组成部分，节日作为人们一种联系情感、传承文化的生活仪式，弘扬节日文化是增强民族凝聚力的重要途径，也是发展民族文化的前提和基础，更是造就和谐社会的强大动力，因此对中华民族传统节日进行研究和挖掘具有重要的意义，能更好地揭示节日文化意义和精神内涵。本文利用情报学知识组织理论和方法对中华民族传统节日进行知识挖掘，是知识组织视域下节日文化基因挖掘的有益探索。

1 节日文化挖掘现状

传统文化对国家的整体发展有着积极的引领作用。近年来，随着我国传统文化保护力度的增加，传统民俗的保护、组织和利用环境得到了改善，不仅建立了丰富的民族文化资源数据库，促进了传统文化事业的发展，也为传统文化价值的挖掘奠定了基础。传统文化知识挖掘逐渐受到学者重视，传统节日的深层知识涉及到节日的时间、人群、地区、起源等多角度的知识提取，每个角度都涉及更深层次的内容，如历史、地理和人文科学，加之节日文化除了存在于各种专业的文献和记录中，更多存在于大众的意识与表达中。因此，与其他类型的非物质文化知识挖掘相比，传统节日文化的知识挖掘具有更多的语义复杂性。

鉴于传统节日文化知识的分散性和多样化，也有部分学者用各种方法来描述和表现传统节日文化中复杂的知识关系。例如，一些学者提出现有的本体论框架可以应用于揭示传统节日文化的隐性知识。在已有的研究成果中，郝挺雷[3]、隗昊等[4]、马旭明等[5]学者提出了为传统节日构建本体库；任廷艳等建议使用XML聚类来挖掘少数民族的节日文化[6]。然而，通过本体挖掘和揭示传统节日文化隐性知识的研究存在两个不足：首先，虽然有可能从本质上揭示和深入挖掘传统节日背后的隐性知识，但是需要大量相关研究人员先围绕知识空间构建本体知识库，然后确定不同主题的知识单元，实现难度和付出成本较大[7]。其次，传统节日文化与大众生活息息相关，文化内容随时间、地域、人群的改变而发生变化，传统节日本体库中的属性值等难以做到及时更新。

因此，本文应用知识组织理论和方法构建传统节日文化标签系统，并挖掘出已标签化的节日文化相关性，深层次揭示传统节日文化资源中隐含的隐性知识。首先，利用知识组织方法构建标签系统来表达传统节日文化；其次，对网络数据进行收集和提取；最后，在一

* 本文系国家社会科学基金项目"图书馆古籍文献的数字人文开发与应用模式研究"（项目编号：17XTQ003）研究成果之一。
作者简介：苏瑞竹（ORCID：0000-0002-5568-4994），副院长，研究馆员；闫静雅（ORCID：0000-0001-6049-9382），硕士研究生；欧阳剑（ORCID：0000-0001-5867-2852），研究馆员，博士，通讯作者，E-mail：oyjjj@163.com。
收稿日期：2019-07-18 修回日期：2019-09-26 本文起止页码：124-132 本文责任编辑：王传清

定数据的支持下,从多个维度对其背后隐性知识进行挖掘。该方法不仅可以充分利用网络上的传统节日文化标签化,在一定程度上克服建立本体知识库所需要的大量人工成本的问题,同时缓解了传统节日文化发展给研究带来的困难。

2 传统节日文化标签系统结构及属性

2.1 传统节日文化标签系统结构及内涵

本文中“传统节日”是指:具有较长历史,能够体现中华民族文化特点的节日,如:春节、端午节等,不涉及现代设定的各类具有商业性质的旅游文化节,如:青岛金沙滩文化旅游节等。

由于节日文化内容的碎片化特点,很多节日文化隐藏于各类典籍、文章或神话传说中,难以提取出其中的隐含关联。因此,本文首先对各传统节日定义多个属性,如:节日日期、节日活动、节日饮食等,以能够全面反映节日文化为目的。其次,对节日文化内容进行提炼,以标签的形式表达,有利于展示节日特征,从而便于发现各属性、标签间的关系,并以此挖掘传统节日文化背后的隐性知识,如图 1 所示:

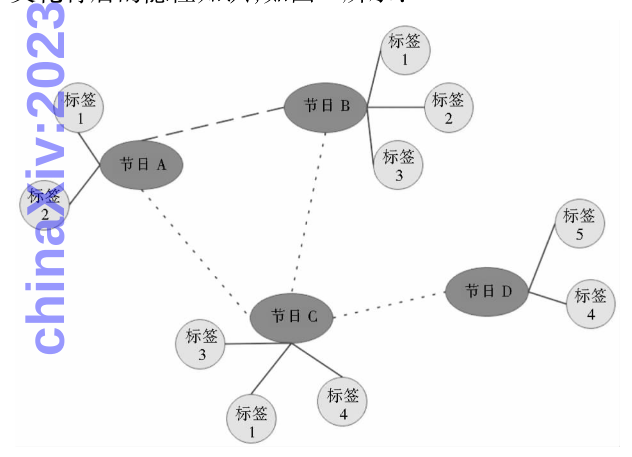


图 1 传统节日文化标签结构

本文将传统节日文化标签系统定义为 S,且 $S = (F_S, T_S, O_S)$,其中 $F_S = \{f_{S1}, f_{S2}, f_{S3}, \cdots, f_{Sn} \mid n \in [1, N]\}$ 是各种传统节日的集合,N 为传统节日的总数量;$T_{Sn} = \{t_{Sn1}, t_{Sn2}, t_{Sn3}, \cdots, t_{Snx} \mid n \in [1, N], x \in [1, X]\}$ 是传统节日文化标签系统中所有标签的集合,是本系统构建的核心内容, X 是对应第 n 个传统节日文化的标签数量;$O_S = \{o_{s1}, o_{s2}, o_{s3}, \cdots, o_{sm} \mid m \in [1, M]\}$ 是传统节日文化标签数据来源的集合,其中 M 为数据源个数。

传统节日文化标签系统中,如果两个传统节日同时被多个相同的标签揭示,那么表明这两个传统节日之间存在着重要的内在关联,或其某些属性可能存在着更深的关联。例如:林王节、杀龙节、茶歌节均为侗族的传统节日,所以这些节日之间存在着相同的“侗族”标签,因此证明它们之间存在着一定的关联;例如壮族、黎族、彝族、瑶族等少数民族会同时在农历三月三日庆祝各族的传统节日,虽然节日名称有所不同,但是节日日期相同,说明这些节日或民族之间存在着一定的关联。依据同样的方法可以得出不同节日、不同民族、不同宗教在饮食习惯、原始信仰、生产方式、历史渊源、族群迁徙、传统文化等方面的内在关联,并且出现相同标签的频数越高,说明两个传统节日或及传统节日属性之间的关联度越大。由此,可以依据不同节日拥有标签的关联程度,归纳和分割出不同民族、不同宗教人群之间的共性和差异。

2.2 传统节日核心属性及层次

本文参考本体工程构建方法之一的七步法对传统节日文化内容进行分析,并以此确定标签属性。七步法是由斯坦福大学开发用于领域本体的构建方法,七个步骤分别是:①确定本体的专业领域和范畴;②考查复用现有本体的可能性;③列出本体中的重要术语;④定义类和类的等级体系;⑤定义类的属性;⑥定义属性的分面;⑦创建实例[8]。根据实际的需要,传统节日标签属性的分析主要分以下 4 个步骤:

(1)确定传统节日文化的专业领域及范畴。构建标签系统,首先确定该标签系统所表达的专业领域及范畴,传统节日文化的专业领域即传统节日在产生和演变过程中,人们将其区别于其他日期的特殊描述、行为以及节日涉及的人群等领域。构建传统节日文化标签系统的目的之一是清晰地表达传统节日文化,因此要确定传统节日文化所涵盖的范围,明确表达的内容,以稳定知识表达的结构。

(2)确定传统节日核心属性的范围。节日属性是对于节日某一方面特征的统称,也是研究和看待节日的不同角度,而节日文化就是这些属性的集合。通过对上一步资料的总结归纳,提出传统节日核心属性:“节日名称”“节日别称”“节日日期”“节日活动”“节日饮食”“相关人物”“节日民族”“节日起源”“起源地点”“起源时间”“神话传说”“流行地区”及“相关宗教”。

(3)明确传统节日核心属性。

节日名称:某一节日由官方机构给出或流传最广的名称。

节日别称:某一节日除“节日名称”以外的名称。

节日日期:节日发生的日期或某一时间段。

节日活动:节日期间大众用以庆祝、纪念的行为和

方法。

节日饮食:节日期间大众食用的特殊食物。

相关人物:与节日有关的真实存在的历史人物。

节日民族:节日期间会进行庆祝的民族。

节日起源:节日的产生原因,如节日纪念的事件、任务或节日由来的信仰文化。

起源地点:节日起源的地区。

起源时间:节日起源的时代或年份。

神话传说:与节日的产生、演变有关的非历史故事。

流行地区:节日现在所流行的区域。

相关宗教:与节日起源或演变有关的宗教。

(4)传统节日核心属性层次划分。由于一些标签在网络上的存在形式为简单的词语或短语,数据提取较为方便,而另一些标签则需要从文章内容中提取,难以自动提取,往往需要辅以人工整理,因此本文将根据标签的提取难度,将属性划分为浅层属性、中层属性、深层属性 3 个等级(见表 1)。

浅层属性指数据源中已标注的、能够直接使用原始数据形成标签的属性;中层属性指需要从文章段落中提取词语或短语形成标签的属性;深层属性指需要对文章或故事的内容进行理解并总结成标签的属性,受数据处理人员主观影响较大。本文以 13 个属性为目标选取 3 个数据源的数据形成标签,构建传统节日文化标签系统模型 S’。

表 1 属性层次划分

属性层次		属性名称和标签示例				
浅层属性	属性	节日名称	节日别称	节日日期	节日民族	
	标签示例	春节、端午节等	新年、过年等	正月初一、五月初五等	汉族、壮族等	
中层属性	属性	节日活动	节日饮食	相关人物	流行地区	
	标签示例	祭祖、对歌等	饺子、糍粑等	屈原、伍子胥等	中国全境、云南边陲等	
深层属性	属性	节日起源	起源地点	起源时间	神话传说	相关宗教
	标签示例	祖先信仰、英雄信仰等	广西地区、中国南部等	上古时代、春秋时期等	年兽传说、莫一大王传说等	佛教、道教等

3 传统节日文化标签系统模型构建

利用标签系统开展知识挖掘这一研究方向,起源于“Folksonomy”这一概念的提出。Folksonomy 意为大众分类法,即允许大众对于信息资源进行分类与标注。在此基础上又出现了“社会化标签”“社会化书签”等概念。社会化标签是我国图书情报学科领域的研究热点,标签被广泛地应用于信息检索、数据可视化、个性化服务、知识挖掘等应用领域。不少图情领域的学者从复杂网络、本体、聚类等角度对标签信息进行了研究,探讨了标签间的关系与深层语义信息[9],发现标签间具有多种关系,如共现关系、上下层级关系、相关关系等,加强对某主题标签间关系的挖掘有利于展现该主题的隐性知识。因此,本文提出将传统节日文化标签化,通过构建传统节日文化标签系统的方式,展现标签间的关系,为开展传统节日文化的研究提供思路。

3.1 构建流程

构建传统节日文化标签系统的第一步是进行属性定义,即定义传统节日文化的核心概念,形成节日文化标签的上层结构。同时,根据研究目的选取数据来源。第二步是从数据来源中提取有关数据,并进行归纳与整合,将整理出的数据形成标签,对应到上层属性之中。由此可以建立传统节日文化标签系统模型 S’,如图 2 所示:

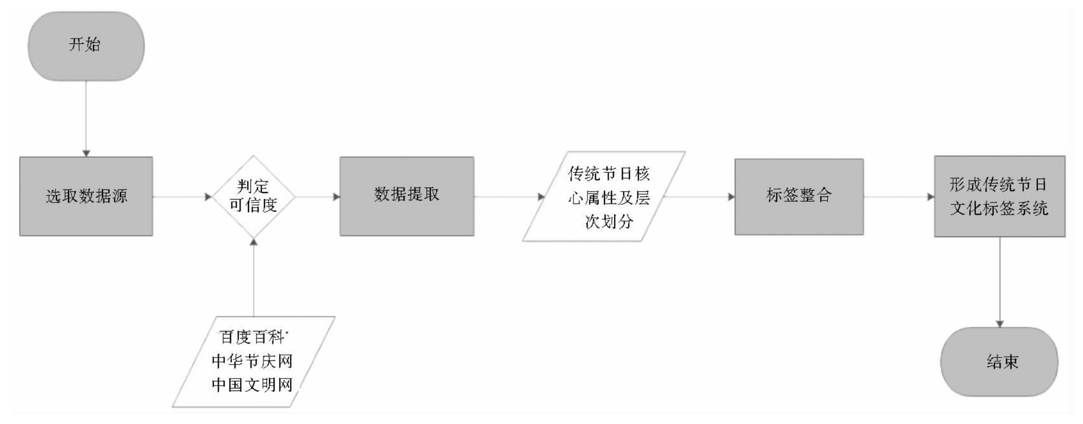


图 2 传统节日文化标签系统模型构建流程

3.2 数据来源与提取

3.2.1 数据来源

传统节日文化存在于官方权威数据库,也存在于百度百科等可以被大众编辑的数据源中。官方权威数据库由传统节日文化领域比较权威的机构出版或搭建,其数据具有科学、准确、可信度高的属性,但数据量较少;百科网站数据量较大且具有一定的可信度。

(1)权威数据库。“中国文明网”是由中共中央宣传部和中央文明办主办[10];“中华节庆网”是由国家民族事物委员会指导,中国人类学民族学研究会民族节庆专业委员会主办[11]。这两个网站都由官方机构托管,数据的权威性和真实性很高。

(2)百度百科。“百度百科”是一部网络百科全书平台,其中包含了海量的高质量半结构化数据,包含了文章标题、分类标签、信息框等大量规范化知识,且其中的词条大部分由人工进行编辑,编辑内容经过审核后方可发布,具有一定可信度。

3.2.2 数据提取

本文采取了两种数据提取策略:策略一,对于网页结构较为清晰的、能够直接获取网络标签的,利用爬虫识别并提取传统节日文化属性及属性值的数据源,本文针对不同数据网站制定不同的抽取模板和抽取规则,使用集搜客爬虫软件,实现网页信息的自动爬取,快速并有效地抽取不同数据源网站的候选数据。首先,需针对样本页面(如:“中国文明网-正月节日”页面内容)建立一个抓取规则(见图3),即通过标注样本页面上所需采集的字段,将抓取规则中的标签与网页信息建立映射关系,再通过添加抓取线索以抓取其他相同结构页面(如:“中国文明网-二月节日”页面内容)上的信息。其次,由于百度百科节日词条网址难以从网页直接获取,因此需要通过关键词列表构造线索并爬取相关页面信息。本文在中国文明网与中华节庆网中共获取整理263个传统节日,以此形成关键词目录。策略二,对于爬虫软件难以自动识别相应属性与属性值的,本文采取人工提取的方法,用人工从文章段落之中提取传统节日属性值,并与属性一一对应,形成对传统节日的描述。

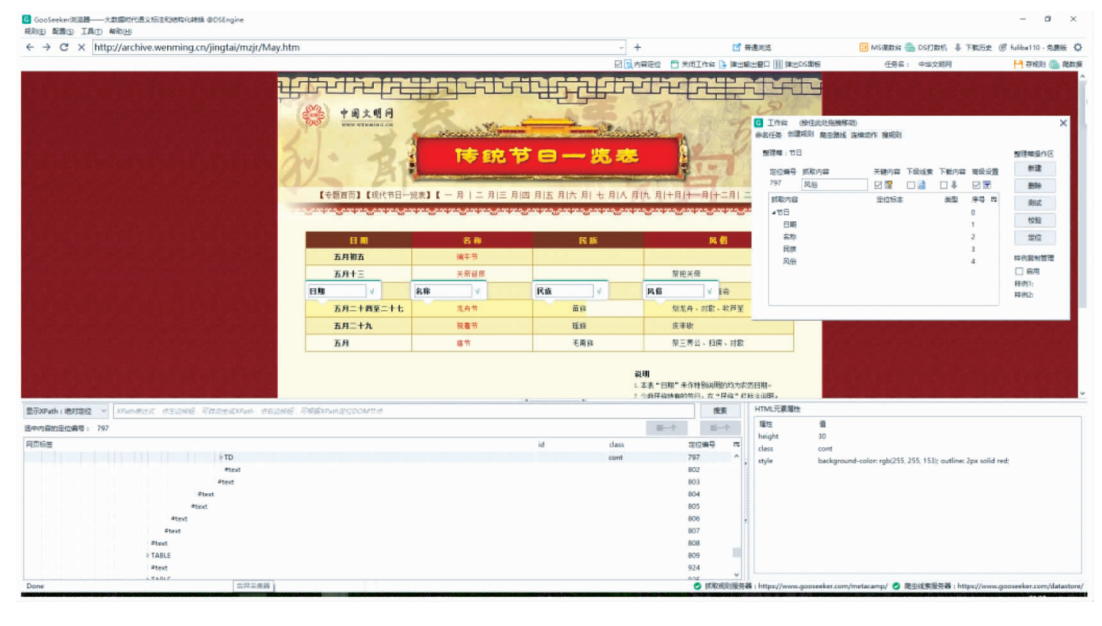


图3 爬取规则设计示意

3.3 标签整合

由于3个数据源的数据均为半结构化数据,且3个数据源在数据结构、语言表述方面存在差异,不同词汇表示同一含义的同义词情况多有出现,例如:节日的“名称”与“中文名”表述不同但含义相同,因此需要进行清洗、去重等工作,减少冗余知识,实现数据整合。

属性对齐则是标签整合的重要内容,属性对齐即异构数据源中属性词汇的统一。由于本文选择的数据源来自3个不同的网站,在部分属性上不同的网站对同一实体的同一属性名使用了不同的描述,而对同一个数据源网站对于固定属性都使用了固定的描述。因此,通过人为构建同义词映射表可以确保不同数据源属性对齐的正确性[12]。例如对于节庆活动的描述,百度百科使用了“节日活动”,而中国文明网使用了“风俗”,因此可以为这些属性建立统一的属性同义词映射表{节日活动,节庆活动、风俗},使这些属性下的标签

统一对应到“节日活动”这个属性中。

标签整合完成后，标签与属性将一一对应，即将标签填充进入各属性中，形成某单一节日的文化标签（见图4），以春节为例：

属性：节日名称；标签：春节；

属性：节日别称；标签：新岁、新年……；

属性：节日日期；标签：正月初一；

属性：节日民族；标签：汉族、满族……；

属性：节日活动；标签：贴年红、拜年、守岁……；

属性：节日饮食；标签：年糕、春卷、饺子……；

……

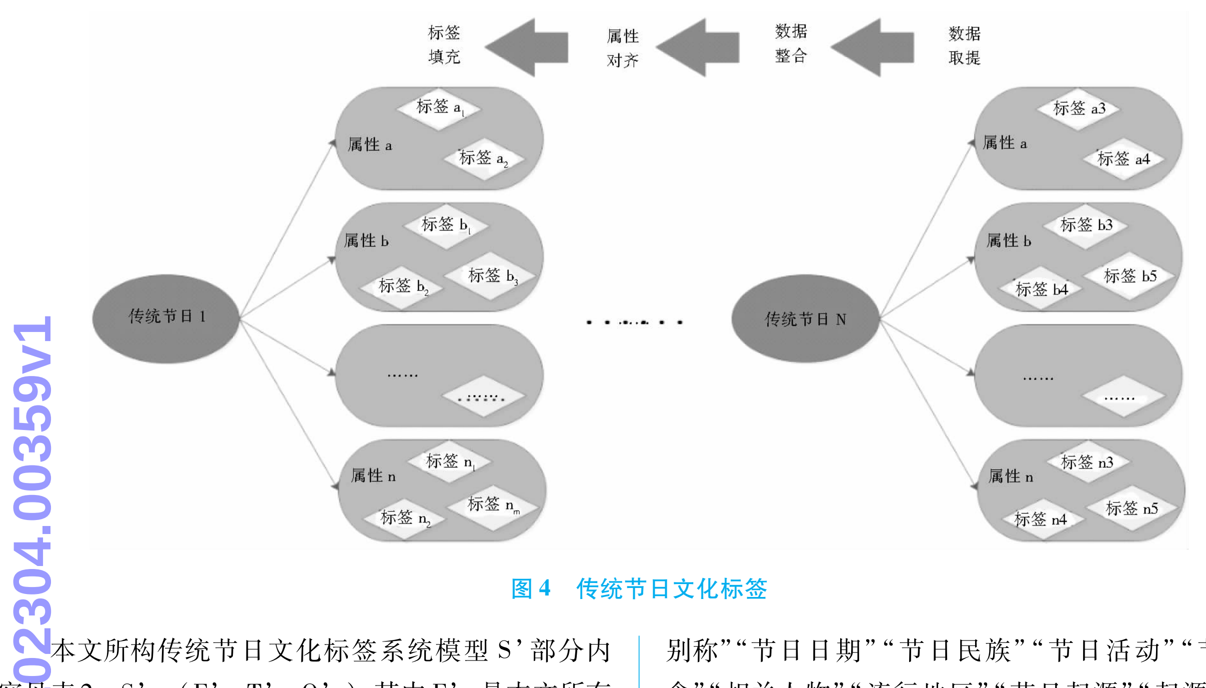


图 4　传统节日文化标签

本文所构传统节日文化标签系统模型 S’部分内容见表2。S’ =（F’$_S$，T’$_S$，O’$_S$），其中 F’$_S$ 是本文所有传统节日的集合，共获取整理传统节日 263 个，N = 263：春节等；T’$_S$ 是本文传统节日文化标签系统中所有标签的集合，分属于 13 个属性：“节日名称”“节日别称”“节日日期”“节日民族”“节日活动”“节日饮食”“相关人物”“流行地区”“节日起源”“起源地点”“起源时间”“神话传说”和“相关宗教”；O’$_S$ 是传统节日文化标签数据来源的集合，M = 3：“中国文明网 - 传统节日”“中华节庆网 - 民族节庆”和百度百科。

表 2　传统节日文化标签系统（部分）

浅层属性				中层属性				深层属性				
节日名称	节日别称	节日日期	节日民族	节日活动	节日饮食	有关人物	流行地区	节日起源	起源地点	起源时间	神话传说	有关宗教
			未分民族	贴年红、守岁、拜年、祭祀、舞狮	茶果点心、年糕、饺子、春卷、汤圆							
春节	新春、新岁、新年等	正月初一	壮族	对歌、打陀螺、跳舞、赛球	扣肉、粉精肉、叉烧肉、白斩鸡、炖猪脚、炖整鸡、粽子	/	中国、东亚、东南亚等	天地信仰、祖先信仰	中国	上古时代	除夕传说、年兽传说	/
			蒙古族	扫除、沐浴、布置蒙古	手把肉							
	……		……	……	……	……	……	……	……	……	……	……
端午节	端阳节、龙日、龙舟节等	五月初五	未分民族	扒龙船、放纸鸢、挂艾草菖蒲	粽子	屈原、伍子胥等	中国及汉字文化圈诸国	天象崇拜、龙图腾祭祀	/	上古时代	/	/
雪顿节	酸奶节、晒佛节、藏戏节	藏历六月二十九日至七月一日	藏族	藏戏演出	酸奶子	/	西藏	僧人解禁出寺下山、老百姓酿酸奶犒劳僧人	西藏拉萨哲蚌寺	公元 11 世纪中叶	/	藏传佛教
……	……	……	……	……	……	……	……	……	……	……	……	……

4 传统节日文化属性层次隐性知识挖掘

本文在传统节日文化标签系统模型 S’的基础上，对其中的标签数据进行统计分析，对传统节日文化属性层次隐性知识展开深入挖掘。在挖掘传统节日文化隐性知识时，不同层次属性的标签数据处理方式也大有不同，浅层属性的标签客观性和稳定性较强，能够在原始数据的基础上直接开展统计分析，但中层或深层属性的标签则需要根据研究需求进行进一步处理。因只涉及挖掘思路的研究，故仅从 3 层属性中选取节日日期、节日民族、节日饮食、节日起源 4 个维度对传统节日文化标签系统进行挖掘实验，并将实验过程与结果进行简单阐述。

4.1 浅层属性隐性知识挖掘

4.1.1 节日日期维度隐性知识挖掘

节日之所以会产生，是因为节日当天或节日前后发生了需要被铭记的事情或变化，于是人们将这个日子赋予与普通日期不同的意义，以此作为纪念，并成为日常生活中的提示，一般每年一个循环。日历是人类最重要的发明之一，人们生活在四季更替、自然变化之中，因此发明了日历以方便记录时间的流逝和四季的变化，而节日则记录了时间长河中最重要的节点。因此，统计节日日期在一年中的分布情况，能够清晰地观察到人们生活中最重要时间段，从中找出相关规律，探寻背后的原因，在知识挖掘过程中，汇总是最简单的一种挖掘方法，浅层属性隐性知识挖掘主要采用数据汇总的方法。

剔除原始数据中节日日期属性下标签数量为 0 或未以农历计算日期的传统节日后，共得到 161 个传统节日，计算每个月份的节日个数并将其导入 Excel 中形成节日数量的月份分布的柱状图(见图 5)。传统节日多数集中在上半年，下半年节日数量较少，正月里传统节日最多，十一月期间节日最少，处在腊月的节日在下半年中相对较多，节日日期的分布规律与农事节气的分布规律相似。正月和腊月由于是年岁更替的前后，腊月里天气寒冷，农事清闲，人们有时间进行各种祭祀、欢聚活动；正月人们在欢庆春节的同时，也开始为新一年的生活生产做出计划，由此衍生出许许多多的节日。所以除了春节这一大传统节日，更产生了许多小节，如：汉族传统的小年(腊月二十三)、苗族的跳坡节(正月初九)。随着春天的到来，天气转暖，万物复苏，人们也开始了农忙的日子。上半年节日较多，首先是为了提醒人们季节的变化，如中和节、侗族的播种节就是为劝民适时耕种、准备进入春耕大忙而设。其次，古时候农事对于人们的生存来说意义非凡，人们往往会进行一些祭祀仪式来祈祷农事顺利，久而久之就产生了节日。

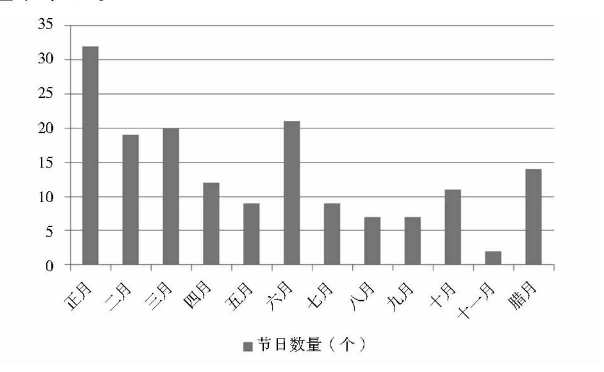


图 5 节日数量的月份分布

4.1.2 节日民族维度隐性知识挖掘

节日是民族身份的重要标志。节日文化蕴含着民族生活中的风土人情、宗教信仰和道德伦理等文化因素，是一个民族历史文化的长期积淀[13]。不同的民族也会庆祝相同的节日，若将“两个民族庆祝相同节日”这一现象定义为“两个民族间存在节日关系”，两个民族共同的节日越多，表明这对民族之间的节日关系越强，那么通过计算民族标签的共现情况就可以观察到民族间节日关系的强弱。

剔除原始数据中民族标签数量为 0 或 1 的数据后，计算两个民族在不同节日中共现的次数，再通过社会网络分析工具 Gephi 对节日民族标签进行分析并可视化(见图 6)。图 6 中节点表示民族标签，节点间的连线表示这对民族在一个或多个节日中存在共现，一个节点的连线数量越多，表示与该民族有节日关系的民族越多，该民族的节日关系网的范围越广；两个节点间的线条越粗，表示这对民族的节日关系越强。由此可见满族、汉族等民族节日关系网范围较广，傣族与德昂族、回族与乌孜别克族等民族间的节日关系较强。

由于地理位置的因素，居住地较近的民族往往节日关系更强。通过观察图 6 可发现民族间的节日关系划分与南北地域划分相似度颇高，南方民族之间、北方民族之间节日关系较强，南方民族与北方民族之间的节日关系较弱，而且发现“北方民族间的节日关系强于南方民族”这一现象，查阅南北民族差异产生原因的相关资料可知，我国南方多是山地丘陵地貌，山河地形阻碍了民族间的相互作用，北方平原面积广袤地形阻碍小。其次，我国北方民族早期多以游牧狩猎为生，人口迁移频繁且面积广泛，而南方民族多是农耕文化，居住

地固定。最后,由于北方少数民族早期兵火较多,各部族相互吞并结合,也造成了现在的民族融合。因此,虽然南北各民族都在历史上有过不少相互作用,但由于地形、农业文明、战乱等因素也造成了一定的南北差异。诸如此类,标签的分布规律能够有效地传达传统节日文化背后的隐性知识,相关研究人员可以通过分析传统节日文化标签进行隐性知识的挖掘工作。

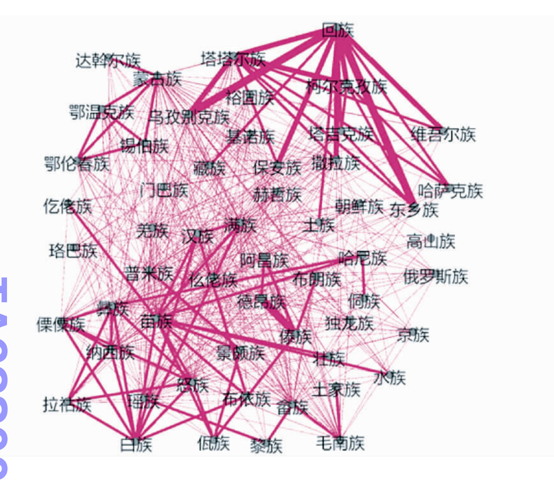


图6 民族间节日关系

4.2 中层属性隐性知识挖掘

中层属性标签是由文章段落中提取词语或短语组成,其表达内容固定,但不同节日的中层属性标签之间往往具有各种不同的关联,而这种关联之间往往隐含着各种难以通过标签本身观察到的隐性知识。以节日饮食标签为例,饮食文化是一个民族重要的文化表征,一个民族的饮食习惯往往和这个民族的农耕文明、生活地区有关。由于人们会在节日当天食用本民族最具特色、最有意义的食物,所以节日饮食更能反映一个民族的文化特征。

通过统计节日饮食在民族间的分布情况,可发现不同民族的饮食特点以及我国的农耕文明。由于一些节日(如:春节)是多民族共同庆祝,因此先对在同一节日中不同民族的饮食标签进行划分,并剔除"鲜果茶点""祭祖金猪""茶果点心"等民族特点不明显的标签,然后将一些同义词合并(如将"油炸馓子/炸馓子"均改为"炸馓子",将"新米/新谷/新米饭"均改为"新米"),得到56个民族的传统节日饮食标签集合,计算饮食标签在民族间的共现情况并使用UCINET软件可进行可视化展示(见图7)。图7中的点表示饮食标签,连线表示这对饮食标签存在共存现象,网络中一个节点的连线越多,关系网越广,那么这个节点也就越重要,其中心度越高,图7中中心度越高点的面积越大。颜色的区分是将节日饮食标签进行聚类,当团体数为20时,是最优分类组合。观察图7,一方面能够从饮食的聚类分布规律中看出民族间的关系:炸油香、炸馓子等食物是回族、哈萨克族等民族热爱的食物,这些民族生活区域接近,且有着相同的宗教文化,酸奶子、切玛等食物是藏族、门巴族等民族较为热衷的事物,同时奶制品也是蒙古等族的节日饮食,这些民族或有着相同的文化信仰或历史居住的环境相似,导致了他们之间相似的饮食习惯。另一方面,从中心度的角度来看,粽子、月饼等饮食中心度较高,说明这些食物为较多的民族所食用。粽子、月饼是我国端午节与中秋节两大传统节日的代表食物,由于我国各民族在历史上的交流与融合,产生了庆祝相同的节日、品尝相同美食的文化。总体来看,由于我国水稻种植面积广、历史长,水稻种植在我国农耕文明的地位决定了稻米在我国各族人民饮食文化中的地位,因此,粽子、糯米糍粑等稻米制作的食物是我国多数民族的传统代表食物。虽然我国以农耕文明为主,但也有小部分的民族拥有独特的游牧民族,如藏、蒙等族,其独特的游牧文化也反映在其节日饮食文化中。

4.3 深层属性隐性知识挖掘

深层属性是从文章内容中归纳而来的,是对节日的基本特征表达与归纳,对深层属性标签的挖掘能够发现节日宏观层面的特征,以节日起源为例,我国的传统节日形式多样,内容丰富,是中华民族悠久历史文化的重要组成部分。从古代到近代,中华民族传统节日的发展过程亦是我国民族历史文化长期积累的过程[14]。我国的传统节日,无一不是从古代甚至远古时期演变而来的。中国古代的这些节日的形成多数因为人们在当时那个科技并不发达的年代对于强大力量的崇拜与信仰,如鬼神传说、宗教信仰、自然节气、英雄人物等。统计节日的起源能够了解古时候人们的原始信仰和思维方式。

由于很多节日起源已经无法查找到具体来历,且往往存在与典籍文章与传奇故事中,文章与故事篇幅长、内容复杂,难以挖掘节日起源包含的隐性知识,因此本文综合牟元奎[15]、苑利等[16]、王娟[17]等学者对于节日产生与来源的理论将节日起源总结为以下6大信仰:鬼神信仰、自然信仰、宗教信仰、英雄信仰、祖先信仰、多信仰结合。通过观察各类起源占比(见图8)可知,没有一种信仰能够占有超过一半的比例,鬼神信仰、英雄信仰、自然信仰在我国古代人民的信仰中占比较大,祖先信仰与宗教信仰占比较小,多信仰结合比例

苏瑞竹，闫静雅，欧阳剑. 传统节日文化标签系统及隐性知识挖掘[J]. 图书情报工作,2020,64(2):124－132.

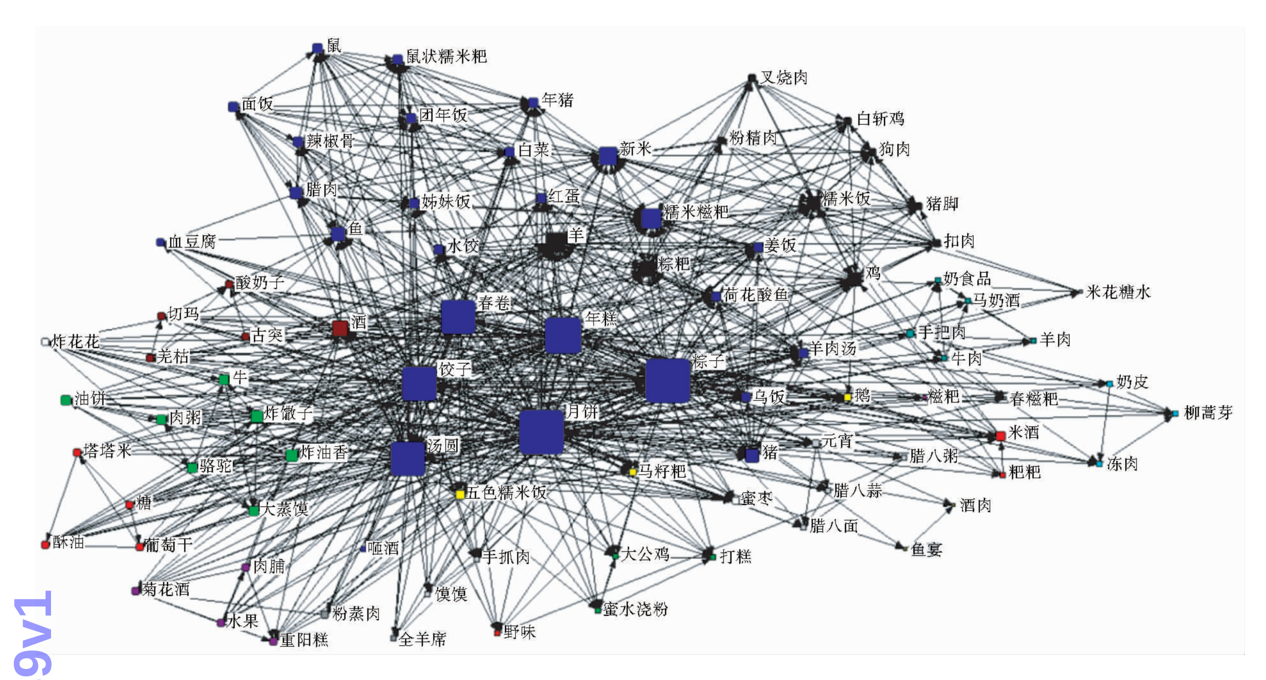


图 7　节日饮食标签共现

较高，其中自然信仰、祖先信仰、鬼神信仰这 3 种起源因素融合较深。在生产不发达的上古时代，当人们无法用科学解释一些自然现象时，便编造一些鬼神故事，企图让这些现象合理化。又因为人们严重依赖自然进行生产生活，在人们眼中这些鬼神的力量变得十分强大，人们不得不采取一些措施讨好鬼神，以求庇护。由于节日起源标签多是关于节日起源的传说或事件，所以大多与某个人物或神怪有关，与祖先关系较小。但节日活动中的“祭祖”这一标签出现 16 次，由此可发现，祖先信仰虽与节日起源关系较小，但对于节日文化的发展产生了很大的影响，人们在重要的日子里往往会选择祭拜祖先。

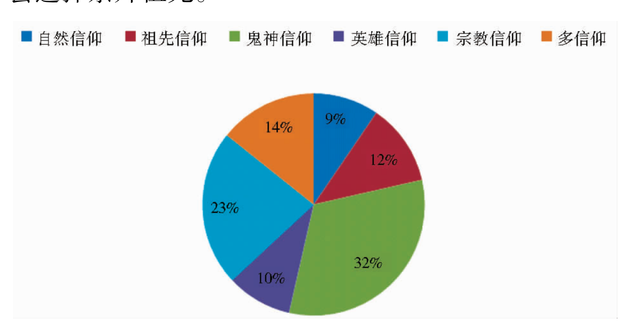


图 8　各类节日起源占比

从单一的属性维度挖掘传统节日文化背后包含的民族间隐性关联，能够从信仰文化、饮食文化等方面深入了解各民族之间的共性与差异。多个属性维度同时挖掘一个主题的内容，能够做到相互支撑，相互验证。再如观察节日民族标签与节日饮食标签的挖掘结果，可以明显看出居住地差异较大的民族在饮食上的差异也较大。单一维度的挖掘难以做到对问题的全面展示，因此多维度的挖掘工作也十分必要。

5　结语

本文从知识组织角度提出传统节日文化标签系统的隐性知识挖掘模式，将传统节日文化标签化的过程，既是对传统节日的解析，又是对文化内容的知识组织与整理。传统节日文化标签系统能够帮助相关的研究人员用标签的形式描述和整理节日文化内容，同时标签的形式也能满足计算机处理的需要，不仅可以充分利用网络上的传统节日文化标签，在一定程度上克服建立本体知识库所需要的大量人工成本的问题，同时缓解了传统节日文化发展给研究带来的困难。知识组织视域下的标签系统这一模式不仅能用在研究传统节日文化方面，更是对我国传统文化的研究有借鉴意义，标签系统为更清晰地整理传统文化提供了框架，为传统文化的表达提供了思路，帮助研究传统文化的学者更直观地观察到传统文化背后的隐性关联。

参考文献：

[1] 萧放. 传统节日：一宗重大的民族文化遗产[J]. 北京师范大学学报(社会科学版),2005(5):50－56.

[2] 文化部外联局国际处. 联合国教科文组织《保护非物质文化遗产公约》基础文件汇编[M]. 北京：外文出版社,2012:7.

[3] 郝挺雷. 传统节日知识本体的构建及表示方法研究[D]. 武汉：华中师范大学,2011.

[4] 隗昊,孙瑜. 基于 OWL 的民族节日领域本体的构建研究[J]. 信息技术与信息化,2016(11):87 - 89.
[5] 马旭明,王海荣. 回族传统节日本体的构建方法研究[J]. 电子技术与软件工程,2018(5):155 - 157.
[6] 任廷艳,罗刚. XML 聚类在少数民族节日文化挖掘中的应用[J]. 软件导刊,2015,14(12):140 - 141.
[7] 程秀峰,毕崇武,李成龙. 基于 SNA 的舞蹈类非物质文化遗产隐性知识关联研究[J]. 图书情报工作,2016,60(2):30 - 36.
[8] 刘仁宁,李禹生. 领域本体构建方法[J]. 武汉工业学院学报,2008(1):46 - 49,53.
[9] 李旭晖,李媛媛,马费成. 我国图情领域社会化标签研究主要问题分析[J]. 图书情报工作,2018,62(16):120 - 131.
[10] 中国文明网编辑部. 中国文明网 - 关于我们[EB/OL]. [2019 - 04 - 02]. http://www.wenming.cn/c/gywm/.
[11] 中国人类学民族学研究会民族节庆专业委员会. 中华节庆网[EB/OL]. [2019 - 04 - 02]. http://www.zhjqw.org/.
[12] 王巍巍,王志刚,潘亮铭,等. 双语影视知识图谱的构建研究[J]. 北京大学学报(自然科学版),2016,52(1):25 - 34.
[13] 覃君英. 少数民族文化资源在《文化生活》教学中的开发与利用研究[D]. 长沙:湖南师范大学,2014.
[14] 罗泽荣. 唯物史观视域下中国传统节日文化研究[D]. 长沙:湖南大学,2009.
[15] 牟元奎. 中国岁时节日的起源与演变[J]. 寻根,1999 (1):9 - 10.
[16] 苑利, 顾军. 中国民俗学教程[M]. 北京:光明日报出版社,2003:101 - 102.
[17] 王娟. 民俗学概论[M]. 北京:北京大学出版社, 2002:15 - 16.

作者贡献说明:

苏瑞竹:提出整体研究思路与框架;论文起草及修改;最终版本修订;

闫静雅:负责数据获取、整理和分析;论文起草及修改;

欧阳剑:负责论文的修改;数据处理技术支撑。

Traditional Festival Culture Tag System and Tacit Knowledge Mining

Su Ruizhu[1] Yan Jingya[1] Ouyang Jian[2]
[1] The School of Management, Guangxi University for Nationalities, Nanning 530006
[2] Library of Shanghai International Studies University, Shanghai 201620

Abstract: [Purpose/significance] The complexity of traditional festival culture expression often makes it difficult to mine more deep meanings. In this research, the tag system is constructed by knowledge organization method to organize and express traditional festival culture knowledge and mine the tacit knowledge behind festival culture. [Method/process] Firstly, this paper designed and defined the structure and attributes of the traditional festival culture tag system. Secondly, it crawled and collated some network data. Finally, the tacit relationship behind it was excavated from multiple dimensions. [Result/conclusion] This paper presents a method of tacit knowledge mining based on traditional festival culture tag system. The experimental results are consistent with the objective law, which proves that it has certain validity and reliability.

Keywords: traditional festival tag system tacit knowledge knowledge mining